GENERATION OF PLANTS WITH ALTERED OIL CONTENT

REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. provisional application 60/530,828 filed December 17, 2003, the contents of which are hereby incorporated by reference.

BACKGROUND OF THE INVENTION

[0002] The ability to manipulate the composition of crop seeds, particularly the content and composition of seed oils, has important applications in the agricultural industries, relating both to processed food oils and to oils for animal feeding. Seeds of agricultural crops contain a variety of valuable constituents, including oil, protein and starch. Industrial processing can separate some or all of these constituents for individual sale in specific applications. For instance, nearly 60% of the U.S. soybean crop is crushed by the soy processing industry. Soy processing yields purified oil, which is sold at high value, while the remainder is sold principally for lower value livestock feed (U.S. Soybean Board, 2001 Soy Stats). Canola seed is crushed to produce oil and the co-product canola meal (Canola Council of Canada). Nearly 20% of the 1999/2000 U.S. corn crop was industrially refined, primarily for production of starch, ethanol and oil (Corn Refiners Association). Thus, it is often desirable to maximize oil content of seeds. For instance, for processed oilseeds such as soy and canola, increasing the absolute oil content of the seed will increase the value of such grains. For processed corn it may be desired to either increase or decrease oil content, depending on utilization of other major constituents. Decreasing oil may improve the quality of isolated starch by reducing undesired flavors associated with oil oxidation. Alternatively, in ethanol production, where flavor is unimportant, increasing oil content may increase overall value. In many fed grains, such as corn and wheat, it is desirable to increase seed oil content, because oil has higher energy content than other seed constituents such as carbohydrate. Oilseed processing, like most grain processing businesses, is a capital-intensive business; thus small shifts in the distribution of products from the low valued components to the high value oil component can have substantial economic impacts for grain processors.

[0003] Biotechnological manipulation of oils can provide compositional alteration and improvement of oil yield. Compositional alterations include high oleic soybean and corn oil (U.S. Pat Nos 6,229,033 and 6,248,939), and laurate-containing seeds (U.S. Pat No 5,639,790), among others. Work in compositional alteration has predominantly focused on processed oilseeds but has been readily extendable to non-oilseed crops, including corn. While there is considerable interest in increasing oil content, the only currently practiced biotechnology in this area is High-Oil Corn (HOC) technology (DuPont, U.S. PAT NO: 5,704,160). HOC employs high oil pollinators developed by classical selection breeding along with elite (male-sterile) hybrid females in a production system referred to as TopCross. The TopCross High Oil system raises harvested grain oil content in maize from about 3.5% to about 7%, improving the energy content of the grain.

[0004] While it has been fruitful, the HOC production system has inherent limitations. First, the system of having a low percentage of pollinators responsible for an entire field's seed set contains inherent risks, particularly in drought years. Second, oil contents in current HOC fields have plateaued at about 9% oil. Finally, high-oil corn is not primarily a biochemical change, but rather an anatomical mutant (increased embryo size) that has the indirect result of increasing oil content. For these reasons, an alternative high oil strategy, particularly one that derives from an altered biochemical output, would be especially valuable.

[0005] The most obvious target crops for the processed oil market are soy and rapeseed, and a large body of commercial work (e.g., U.S. Pat No: 5,952,544; PCT application WO9411516) demonstrates that *Arabidopsis* is an excellent model for oil metabolism in these crops. Biochemical screens of seed oil composition have identified *Arabidopsis* genes for many critical biosynthetic enzymes and have led to identification of agronomically important gene orthologs. For instance, screens using chemically mutagenized populations have identified lipid mutants whose seeds display altered fatty acid composition (Lemieux et al., 1990; James and Dooner, 1990). T-DNA mutagenesis screens (Feldmann et al., 1989) that detected altered fatty acid composition identified the omega 3 desaturase (*FAD3*) and delta-12 desaturase (*FAD2*) genes (U.S. Pat No 5952544; Yadav et al., 1993; Okuley et al., 1994). A screen which focused on oil content rather than oil quality, analyzed chemically-induced mutants for wrinkled seeds or altered seed density, from which altered seed oil content was inferred (Focks and Benning, 1998). Another screen, designed to identify enzymes involved in production of very long chain fatty acids, identified a mutation in the gene encoding a diacylglycerol acyltransferase (DGAT) as being responsible for reduced triacyl glycerol accumulation in seeds (Katavic V et al, 1995). It was further shown that seed-specific over-expression of the DGAT cDNA was associated with increased seed oil content (Jako et al., 2001).

[0006] Activation tagging in plants refers to a method of generating random mutations by insertion of a heterologous nucleic acid construct comprising regulatory sequences (e.g., an enhancer) into a plant genome. The regulatory sequences can act to enhance transcription of one or more native plant genes; accordingly, activation tagging is a fruitful method for generating gain-of-function, generally dominant mutants (see, *e.g.*, Hayashi *et al.*, 1992; Weigel D *et al.* 2000). The inserted construct provides a molecular tag for rapid identification of the native plant whose mis-expression causes the mutant phenotype. Activation tagging may also cause loss-of-function phenotypes. The insertion may result in disruption of a native plant gene, in which case the phenotype is generally recessive.

[0007] Activation tagging has been used in various species, including tobacco and *Arabidopsis*, to identify many different kinds of mutant phenotypes and the genes associated with these phenotypes (Wilson *et al.*, 1996, Schaffer *et al.*, 1998, Fridborg *et al.*, 1999; Kardailsky *et al.*, 1999; Christensen S *et al.* 1998).

SUMMARY OF THE INVENTION

[0008] The invention provides a transgenic plant having a high oil phenotype. The transgenic plant comprises a transformation vector comprising a nucleotide sequence that encodes or is complementary to a sequence that encodes a HIO30.5 polypeptide comprising the amino acid sequence of SEQ ID NO:2, or an ortholog thereof. In preferred embodiments, the transgenic plant is selected from the group consisting of rapeseed, soy, corn, sunflower, cotton, cocoa, safflower, oil palm, coconut palm, flax, castor and peanut. The invention also provides a method of producing oil comprising growing the transgenic plant and recovering oil from said plant. The invention further provides a method of generating a plant having a high oil phenotype by identifying a plant that has an allele in its HIO30.4 gene that results in increased oil content compared to plants lacking the allele and generating progeny of the identified plant, wherein the generated progeny inherit the allele and have the high oil phenotype.

[0009] The transgenic plant of the invention is produced by a method that comprises introducing into progenitor cells of the plant a plant transformation vector comprising a nucleotide sequence that encodes or is complementary to a sequence that encodes a HIO30.5 polypeptide, and growing the transformed progenitor cells to produce a transgenic plant, wherein the HIO30.5 polynucleotide sequence is expressed causing the high oil phenotype.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

[0010] Unless otherwise indicated, all technical and scientific terms used herein have the same meaning as they would to one skilled in the art of the present invention. Practitioners are particularly directed to Sambrook *et al.*, 1989, and Ausubel FM *et al.*, 1993, for definitions and terms of the art. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary.

[0011] As used herein, the term "vector" refers to a nucleic acid construct designed for transfer between different host cells. An "expression vector" refers to a vector that has the ability to incorporate and express heterologous DNA fragments in a foreign cell. Many

prokaryotic and eukaryotic expression vectors are commercially available. Selection of appropriate expression vectors is within the knowledge of those having skill in the art.

[0012] A "heterologous" nucleic acid construct or sequence has a portion of the sequence that is not native to the plant cell in which it is expressed. Heterologous, with respect to a control sequence refers to a control sequence (*i.e.* promoter or enhancer) that does not function in nature to regulate the same gene the expression of which it is currently regulating. Generally, heterologous nucleic acid sequences are not endogenous to the cell or part of the genome in which they are present, and have been added to the cell, by infection, transfection, microinjection, electroporation, or the like. A "heterologous" nucleic acid construct may contain a control sequence/DNA coding sequence combination that is the same as, or different from a control sequence/DNA coding sequence combination found in the native plant.

[0013] As used herein, the term "gene" means the segment of DNA involved in producing a polypeptide chain, which may or may not include regions preceding and following the coding region, *e.g.* 5' untranslated (5' UTR) or "leader" sequences and 3' UTR or "trailer" sequences, as well as intervening sequences (introns) between individual coding segments (exons) and non-transcribed regulatory sequence.

[0014] As used herein, "recombinant" includes reference to a cell or vector, that has been modified by the introduction of a heterologous nucleic acid sequence or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found in identical form within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all as a result of deliberate human intervention.

[0015] As used herein, the term "gene expression" refers to the process by which a polypeptide is produced based on the nucleic acid sequence of a gene. The process includes both transcription and translation; accordingly, "expression" may refer to either a polynucleotide or polypeptide sequence, or both. Sometimes, expression of a polynucleotide sequence will not lead to protein translation. "Over-expression" refers to increased expression of a polynucleotide and/or polypeptide sequence relative to its expression in a wild-type (or other reference [e.g., non-transgenic]) plant and may relate to a naturally-occurring or non-naturally occurring sequence. "Ectopic expression" refers to expression at a time, place, and/or increased level that does not naturally occur in the non-altered or wild-type plant. "Under-expression" refers to decreased expression of a polynucleotide and/or polypeptide sequence, generally of an endogenous gene, relative to

its expression in a wild-type plant. The terms "mis-expression" and "altered expression" encompass over-expression, under-expression, and ectopic expression.

[0016] The term "introduced" in the context of inserting a nucleic acid sequence into a cell, means "transfection", or "transformation" or "transduction" and includes reference to the incorporation of a nucleic acid sequence into a eukaryotic or prokaryotic cell where the nucleic acid sequence may be incorporated into the genome of the cell (for example, chromosome, plasmid, plastid, or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (for example, transfected mRNA).

[0017] As used herein, a "plant cell" refers to any cell derived from a plant, including cells from undifferentiated tissue (*e.g.*, callus) as well as plant seeds, pollen, progagules and embryos.

[0018] As used herein, the terms "native" and "wild-type" relative to a given plant trait or phenotype refers to the form in which that trait or phenotype is found in the same variety of plant in nature.

[0019] As used herein, the term "modified" regarding a plant trait, refers to a change in the phenotype of a transgenic plant relative to the similar non-transgenic plant. An "interesting phenotype (trait)" with reference to a transgenic plant refers to an observable or measurable phenotype demonstrated by a T1 and/or subsequent generation plant, which is not displayed by the corresponding non-transgenic (i.e., a genotypically similar plant that has been raised or assayed under similar conditions). An interesting phenotype may represent an improvement in the plant or may provide a means to produce improvements in other plants. An "improvement" is a feature that may enhance the utility of a plant species or variety by providing the plant with a unique and/or novel quality. An "altered oil content phenotype" refers to measurable phenotype of a genetically modified plant, where the plant displays a statistically significant increase or decrease in overall oil content (i.e., the percentage of seed mass that is oil), as compared to the similar, but non-modified plant. A high oil phenotype refers to an increase in overall oil content.

[0020] As used herein, a "mutant" polynucleotide sequence or gene differs from the corresponding wild type polynucleotide sequence or gene either in terms of sequence or expression, where the difference contributes to a modified plant phenotype or trait. Relative to a plant or plant line, the term "mutant" refers to a plant or plant line which has a modified plant phenotype or trait, where the modified phenotype or trait is associated with the modified expression of a wild type polynucleotide sequence or gene.

[0021] As used herein, the term "T1" refers to the generation of plants from the seed of T0 plants. The T1 generation is the first set of transformed plants that can be selected by application of a selection agent, *e.g.*, an antibiotic or herbicide, for which the transgenic plant contains the corresponding resistance gene. The term "T2" refers to the generation of plants by self-fertilization of the flowers of T1 plants, previously selected as being transgenic. T3 plants are generated from T2 plants, etc. As used herein, the "direct progeny" of a given plant derives from the seed (or, sometimes, other tissue) of that plant and is in the immediately subsequent generation; for instance, for a given lineage, a T2 plant is the direct progeny of a T1 plant. The "indirect progeny" of a given plant derives from the seed (or other tissue) of the direct progeny of that plant, or from the seed (or other tissue) of subsequent generations in that lineage; for instance, a T3 plant is the indirect progeny of a T1 plant.

[0022] As used herein, the term "plant part" includes any plant organ or tissue, including, without limitation, seeds, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, and microspores. Plant cells can be obtained from any plant organ or tissue and cultures prepared therefrom. The class of plants which can be used in the methods of the present invention is generally as broad as the class of higher plants amenable to transformation techniques, including both monocotyledenous and dicotyledenous plants.

[0023] As used herein, "transgenic plant" includes a plant that comprises within its genome a heterologous polynucleotide. The heterologous polynucleotide can be either stably integrated into the genome, or can be extra-chromosomal. Preferably, the polynucleotide of the present invention is stably integrated into the genome such that the polynucleotide is passed on to successive generations. A plant cell, tissue, organ, or plant into which the heterologous polynucleotides have been introduced is considered "transformed", "transfected", or "transgenic". Direct and indirect progeny of transformed plants or plant cells that also contain the heterologous polynucleotide are also considered transgenic.

Identification of Plants with an Altered Oil Content Phenotype

[0024] We used an *Arabidopsis* activation tagging (ACTTAG) screen to identify the association between the gene we have designated "HIO30.5," (At3g52280; GI#18409525:1-1110), and an altered oil content phenotype (specifically, a high oil phenotype). Briefly, and as further described in the Examples, a large number of *Arabidopsis* plants were mutated by transformation with the pSKI015 vector, which comprises a T-DNA from the Ti plasmid of *Agrobacterium tumifaciens*, a viral enhancer

element, and a selectable marker gene (Weigel *et al*, 2000). When the T-DNA inserts into the genome of transformed plants, the enhancer element can cause up-regulation genes in the vicinity, generally within about 10 kilobase (kb) of the enhancers. To identify transgenic plants, T1 plants were exposed to the selective agent in order to specifically recover transformed plants that expressed the selectable marker and harbored the T-DNA. T2 seed was harvested from these plants. Lipids were extracted from of about 15-20 T2 seeds. Gas chromatography (GC) analysis was performed to determine fatty acid content and composition of seed samples.

[0025] An *Arabidopsis* line that showed a high-oil phenotype was identified, wherein oil content (i.e., fatty acids) constituted about 35% of seed mass compared to an average oil content of about 28.7% for seed from other plants grown and analyzed at the same time (a 22% increase in oil). The association of the HIO30.5 gene with the high oil phenotype was discovered by identifying the site of T-DNA insertion and, as shown in the Examples, demonstrating genetic co-segregation of the high seed oil phenotype and the presence of the T-DNA. Accordingly, HIO30.5 genes and/or polypeptides may be employed in the development of genetically modified plants having a modified oil content phenotype ("a HIO30.5 phenotype"). HIO30.5 genes may be used in the generation of oilseed crops that provide improved oil yield from oilseed processing and in the generation of feed grain crops that provide increased energy for animal feeding. HIO30.5 genes may further be used to increase the oil content of specialty oil crops, in order to augment yield of desired unusual fatty acids. Transgenic plants that have been genetically modified to express HIO30.5 can be used in the production of oil, wherein the transgenic plants are grown, and oil is obtained from plant parts (e.g. seed) using standard methods.

HIO30.5 Nucleic Acids and Polypeptides

[0026] *Arabidopsis* HIO30.5 nucleic acid (genomic DNA) sequence is provided in SEQ ID NO:1 and in Genbank entry GI#18409525:1-1110. The corresponding protein sequence is provided in SEQ ID NO:2 and in GI#15231174. Nucleic acids and/or proteins that are orthologs or paralogs of *Arabidopsis* HIO30.5, are described in Example 3 below.

[0027] As used herein, the term "HIO30.5 polypeptide" refers to a full-length HIO30.5 protein or a fragment, derivative (variant), or ortholog thereof that is "functionally active," meaning that the protein fragment, derivative, or ortholog exhibits one or more or the functional activities associated with the polypeptide of SEQ ID NO:2. In one preferred embodiment, a functionally active HIO30.5 polypeptide causes an altered oil content phenotype when mis-expressed in a plant. In a further preferred embodiment, mis-expression of the HIO30.5 polypeptide causes a high oil phenotype in a plant. In another embodiment, a functionally active HIO30.5 polypeptide is capable of rescuing defective

(including deficient) endogenous HIO30.5 activity when expressed in a plant or in plant cells; the rescuing polypeptide may be from the same or from a different species as that with defective activity. In another embodiment, a functionally active fragment of a full length HIO30.5 polypeptide (i.e., a native polypeptide having the sequence of SEQ ID NO:2 or a naturally occurring ortholog thereof) retains one of more of the biological properties associated with the full-length HIO30.5 polypeptide, such as signaling activity, binding activity, catalytic activity, or cellular or extra-cellular localizing activity. A HIO30.5 fragment preferably comprises a beta subunit of transcription initiation factor IIF domain HIO30.5 domain, such as a C- or N-terminal or catalytic domain, among others, and preferably comprises at least 10, preferably at least 20, more preferably at least 25, and most preferably at least 50 contiguous amino acids of a HIO30.5 protein. Functional domains can be identified using the PFAM program (Bateman A et al., 1999 Nucleic Acids Res 27:260-262). A preferred HIO30.5 fragment comprises a bromodomain (PFam00439). Functionally active variants of full-length HIO30.5 polypeptides or fragments thereof include polypeptides with amino acid insertions, deletions, or substitutions that retain one of more of the biological properties associated with the full-length HIO30.5 polypeptide. In some cases, variants are generated that change the post-translational processing of a HIO30.5 polypeptide. For instance, variants may have altered protein transport or protein localization characteristics or altered protein half-life compared to the native polypeptide.

[0028] As used herein, the term "HIO30.5 nucleic acid" encompasses nucleic acids with the sequence provided in or complementary to the sequence provided in SEQ ID NO:1, as well as functionally active fragments, derivatives, or orthologs thereof. A HIO30.5 nucleic acid of this invention may be DNA, derived from genomic DNA or cDNA, or RNA.

[0029] In one embodiment, a functionally active HIO30.5 nucleic acid encodes or is complementary to a nucleic acid that encodes a functionally active HIO30.5 polypeptide. Included within this definition is genomic DNA that serves as a template for a primary RNA transcript (i.e., an mRNA precursor) that requires processing, such as splicing, before encoding the functionally active HIO30.5 polypeptide. A HIO30.5 nucleic acid can include other non-coding sequences, which may or may not be transcribed; such sequences include 5' and 3' UTRs, polyadenylation signals and regulatory sequences that control gene expression, among others, as are known in the art. Some polypeptides require processing events, such as proteolytic cleavage, covalent modification, etc., in order to become fully active. Accordingly, functionally active nucleic acids may encode the mature or the pre-processed HIO30.5 polypeptide, or an intermediate form. A HIO30.5 polynucleotide can also include heterologous coding sequences, for example, sequences

that encode a marker included to facilitate the purification of the fused polypeptide, or a transformation marker.

[0030] In another embodiment, a functionally active HIO30.5 nucleic acid is capable of being used in the generation of loss-of-function HIO30.5 phenotypes, for instance, via antisense suppression, co-suppression, etc.

[0031] In one preferred embodiment, a HIO30.5 nucleic acid used in the methods of this invention comprises a nucleic acid sequence that encodes or is complementary to a sequence that encodes a HIO30.5 polypeptide having at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95% or more sequence identity to the polypeptide sequence presented in SEQ ID NO:2.

[0032] In another embodiment a HIO30.5 polypeptide of the invention comprises a polypeptide sequence with at least 50% or 60% identity to the HIO30.5 polypeptide sequence of SEQ ID NO:2, and may have at least 70%, 80%, 85%, 90% or 95% or more sequence identity to the HIO30.5 polypeptide sequence of SEQ ID NO:2. In another embodiment, a HIO30.5 polypeptide comprises a polypeptide sequence with at least 50%, 60%, 70%, 80%, 85%, 90% or 95% or more sequence identity to a functionally active fragment of the polypeptide presented in SEQ ID NO:2, such as a bromodomain. In yet another embodiment, a HIO30.5 polypeptide comprises a polypeptide sequence with at least 50%, 60 %, 70%, 80%, or 90% identity to the polypeptide sequence of SEQ ID NO:2 over its entire length and comprises a bromodomain.

[0033] In another aspect, a HIO30.5 polynucleotide sequence is at least 50% to 60% identical over its entire length to the HIO30.5 nucleic acid sequence presented as SEQ ID NO:1, or nucleic acid sequences that are complementary to such a HIO30.5 sequence, and may comprise at least 70%, 80%, 85%, 90% or 95% or more sequence identity to the HIO30.5 sequence presented as SEQ ID NO:1 or a functionally active fragment thereof, or complementary sequences.

[0034] As used herein, "percent (%) sequence identity" with respect to a specified subject sequence, or a specified portion thereof, is defined as the percentage of nucleotides or amino acids in the candidate derivative sequence identical with the nucleotides or amino acids in the subject sequence (or specified portion thereof), after aligning the sequences and introducing gaps, if necessary to achieve the maximum percent sequence identity, as generated by the program WU-BLAST-2.0a19 (Altschul *et al.*, J. Mol. Biol. (1990) 215:403-410) with search parameters set to default values. The HSP S and HSP S2 parameters are dynamic values and are established by the program itself depending upon

the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched. A "% identity value" is determined by the number of matching identical nucleotides or amino acids divided by the sequence length for which the percent identity is being reported. "Percent (%) amino acid sequence similarity" is determined by doing the same calculation as for determining % amino acid sequence identity, but including conservative amino acid substitutions in addition to identical amino acids in the computation. A conservative amino acid substitution is one in which an amino acid is substituted for another amino acid having similar properties such that the folding or activity of the protein is not significantly affected. Aromatic amino acids that can be substituted for each other are phenylalanine, tryptophan, and tyrosine; interchangeable hydrophobic amino acids are leucine, isoleucine, methionine, and valine; interchangeable polar amino acids are glutamine and asparagine; interchangeable basic amino acids are arginine, lysine and histidine; interchangeable acidic amino acids are aspartic acid and glutamic acid; and interchangeable small amino acids are alanine, serine, threonine, cysteine and glycine.

[0035] Derivative nucleic acid molecules of the subject nucleic acid molecules include sequences that selectively hybridize to the nucleic acid sequence of SEQ ID NO:1. The stringency of hybridization can be controlled by temperature, ionic strength, pH, and the presence of denaturing agents such as formamide during hybridization and washing. Conditions routinely used are well known (see, *e.g.*, Current Protocol in Molecular Biology, Vol. 1, Chap. 2.10, John Wiley & Sons, Publishers (1994); Sambrook *et al.*, Molecular Cloning, Cold Spring Harbor (1989)). In some embodiments, a nucleic acid molecule of the invention is capable of hybridizing to a nucleic acid molecule containing the nucleotide sequence of SEQ ID NO:1 under stringent hybridization conditions that are: prehybridization of filters containing nucleic acid for 8 hours to overnight at 65° C in a solution comprising 6X single strength citrate (SSC) (1X SSC is 0.15 M NaCl, 0.015 M Na citrate; pH 7.0), 5X Denhardt's solution, 0.05% sodium pyrophosphate and 100 μg/ml herring sperm DNA; hybridization for 18-20 hours at 65° C in a solution containing 6X SSC, 1X Denhardt's solution, 100 μg/ml yeast tRNA and 0.05% sodium pyrophosphate; and washing of filters at 65° C for 1 h in a solution containing 0.1X SSC and 0.1% SDS (sodium dodecyl sulfate). In other embodiments, moderately stringent hybridization conditions are used that are: pretreatment of filters containing nucleic acid for 6 h at 40° C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.1% PVP, 0.1% Ficoll, 1% BSA, and 500 μg/ml denatured salmon sperm DNA; hybridization for 18-20 h at 40° C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 μg/ml salmon sperm DNA, and 10% (wt/vol) dextran sulfate; followed by washing twice for 1 hour at 55° C in a solution containing 2X SSC and 0.1% SDS. Alternatively, low

stringency conditions can be used that comprise: incubation for 8 hours to overnight at 37° C in a solution comprising 20% formamide, 5 x SSC, 50 mM sodium phosphate (pH 7.6), 5X Denhardt's solution, 10% dextran sulfate, and 20 μg/ml denatured sheared salmon sperm DNA; hybridization in the same buffer for 18 to 20 hours; and washing of filters in 1 x SSC at about 37° C for 1 hour.

[0036] As a result of the degeneracy of the genetic code, a number of polynucleotide sequences encoding a HIO30.5 polypeptide can be produced. For example, codons may be selected to increase the rate at which expression of the polypeptide occurs in a particular host species, in accordance with the optimum codon usage dictated by the particular host organism (see, e.g., Nakamura et al, 1999). Such sequence variants may be used in the methods of this invention.

[0037] The methods of the invention may use orthologs of the *Arabidopsis* HIO30.5. Methods of identifying the orthologs in other plant species are known in the art. Normally, orthologs in different species retain the same function, due to presence of one or more protein motifs and/or 3-dimensional structures. In evolution, when a gene duplication event follows speciation, a single gene in one species, such as *Arabidopsis*, may correspond to multiple genes (paralogs) in another. As used herein, the term "orthologs" encompasses paralogs. When sequence data is available for a particular plant species, orthologs are generally identified by sequence homology analysis, such as BLAST analysis, usually using protein bait sequences. Sequences are assigned as a potential ortholog if the best hit sequence from the forward BLAST result retrieves the original query sequence in the reverse BLAST (Huynen MA and Bork P, Proc Natl Acad Sci (1998) 95:5849-5856; Huynen MA *et al.*, Genome Research (2000) 10:1204-1210). Programs for multiple sequence alignment, such as CLUSTAL (Thompson JD et al, 1994, Nucleic Acids Res 22:4673-4680) may be used to highlight conserved regions and/or residues of orthologous proteins and to generate phylogenetic trees. In a phylogenetic tree representing multiple homologous sequences from diverse species (e.g., retrieved through BLAST analysis), orthologous sequences from two species generally appear closest on the tree with respect to all other sequences from these two species. Structural threading or other analysis of protein folding (e.g., using software by ProCeryon, Biosciences, Salzburg, Austria) may also identify potential orthologs. Nucleic acid hybridization methods may also be used to find orthologous genes and are preferred when sequence data are not available. Degenerate PCR and screening of cDNA or genomic DNA libraries are common methods for finding related gene sequences and are well known in the art (see, e.g., Sambrook, 1989; Dieffenbach and Dveksler, 1995). For instance, methods for generating a cDNA library from the plant species of interest and probing the library with partially homologous gene probes are described in Sambrook *et al.* A highly conserved

portion of the *Arabidopsis* HIO30.5 coding sequence may be used as a probe. HIO30.5 ortholog nucleic acids may hybridize to the nucleic acid of SEQ ID NO:1 under high, moderate, or low stringency conditions. After amplification or isolation of a segment of a putative ortholog, that segment may be cloned and sequenced by standard techniques and utilized as a probe to isolate a complete cDNA or genomic clone. Alternatively, it is possible to initiate an EST project to generate a database of sequence information for the plant species of interest. In another approach, antibodies that specifically bind known HIO30.5 polypeptides are used for ortholog isolation (see, e.g., Harlow and Lane, 1988, 1999). Western blot analysis can determine that a HIO30.5 ortholog (i.e., an orthologous protein) is present in a crude extract of a particular plant species. When reactivity is observed, the sequence encoding the candidate ortholog may be isolated by screening expression libraries representing the particular plant species. Expression libraries can be constructed in a variety of commercially available vectors, including lambda gt11, as described in Sambrook, *et al.*, 1989. Once the candidate ortholog(s) are identified by any of these means, candidate orthologous sequence are used as bait (the "query") for the reverse BLAST against sequences from *Arabidopsis* or other species in which HIO30.5 nucleic acid and/or polypeptide sequences have been identified.

[0038] HIO30.5 nucleic acids and polypeptides may be obtained using any available method. For instance, techniques for isolating cDNA or genomic DNA sequences of interest by screening DNA libraries or by using polymerase chain reaction (PCR), as previously described, are well known in the art. Alternatively, nucleic acid sequence may be synthesized. Any known method, such as site directed mutagenesis (Kunkel TA *et al.*, 1991), may be used to introduce desired changes into a cloned nucleic acid.

[0039] In general, the methods of the invention involve incorporating the desired form of the HIO30.5 nucleic acid into a plant expression vector for transformation of in plant cells, and the HIO30.5 polypeptide is expressed in the host plant.

[0040] An isolated HIO30.5 nucleic acid molecule is other than in the form or setting in which it is found in nature and is identified and separated from least one contaminant nucleic acid molecule with which it is ordinarily associated in the natural source of the HIO30.5 nucleic acid. However, an isolated HIO30.5 nucleic acid molecule includes HIO30.5 nucleic acid molecules contained in cells that ordinarily express HIO30.5 where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

Generation of Genetically Modified Plants with an Altered Oil Content Phenotype

[0041] HIO30.5 nucleic acids and polypeptides may be used in the generation of genetically modified plants having a modified oil content phenotype. As used herein, a "modified oil content phenotype" may refer to modified oil content in any part of the plant; the modified oil content is often observed in seeds. In a preferred embodiment, altered expression of the HIO30.5 gene in a plant is used to generate plants with a high oil phenotype.

[0042] The methods described herein are generally applicable to all plants. Although activation tagging and gene identification is carried out in *Arabidopsis*, the HIO30.5 gene (or an ortholog, variant or fragment thereof) may be expressed in any type of plant. In a preferred embodiment, the invention is directed to oil-producing plants, which produce and store triacylglycerol in specific organs, primarily in seeds. Such species include soybean (*Glycine max*), rapeseed and canola (including *Brassica napus*, *B. campestris*), sunflower (*Helianthus annus*), cotton (*Gossypium hirsutum*), corn (*Zea mays*), cocoa (*Theobroma cacao*), safflower (*Carthamus tinctorius*), oil palm (*Elaeis guineensis*), coconut palm (*Cocos nucifera*), flax (*Linum usitatissimum*), castor (*Ricinus communis*) and peanut (*Arachis hypogaea*). The invention may also be directed to fruit- and vegetable-bearing plants, grain-producing plants, nut-producing plants, rapid cycling *Brassica* species, alfalfa (Medicago sativa), tobacco (Nicotiana), turfgrass (Poaceae family), other forage crops, and wild species that may be a source of unique fatty acids.

[0043] The skilled artisan will recognize that a wide variety of transformation techniques exist in the art, and new techniques are continually becoming available. Any technique that is suitable for the target host plant can be employed within the scope of the present invention. For example, the constructs can be introduced in a variety of forms including, but not limited to as a strand of DNA, in a plasmid, or in an artificial chromosome. The introduction of the constructs into the target plant cells can be accomplished by a variety of techniques, including, but not limited to *Agrobacterium*-mediated transformation, electroporation, microinjection, microprojectile bombardment calcium-phosphate-DNA co-precipitation or liposome-mediated transformation of a heterologous nucleic acid. The transformation of the plant is preferably permanent, *i.e.* by integration of the introduced expression constructs into the host plant genome, so that the introduced constructs are passed onto successive plant generations. Depending upon the intended use, a heterologous nucleic acid construct comprising an HIO30.5 polynucleotide may encode the entire protein or a biologically active portion thereof.

[0044] In one embodiment, binary Ti-based vector systems may be used to transfer polynucleotides. Standard *Agrobacterium* binary vectors are known to those of skill in the art, and many are commercially available (e.g., pBI121 Clontech Laboratories, Palo Alto, CA).

[0045] The optimal procedure for transformation of plants with *Agrobacterium* vectors will vary with the type of plant being transformed. Exemplary methods for *Agrobacterium*-mediated transformation include transformation of explants of hypocotyl, shoot tip, stem or leaf tissue, derived from sterile seedlings and/or plantlets. Such transformed plants may be reproduced sexually, or by cell or tissue culture. *Agrobacterium* transformation has been previously described for a large number of different types of plants and methods for such transformation may be found in the scientific literature. Of particular relevance are methods to transform commercially important crops, such as rapeseed (De Block et al., 1989), sunflower (Everett et al., 1987), and soybean (Christou et al., 1989; Kline et al., 1987).

[0046] Expression (including transcription and translation) of HIO30.5 may be regulated with respect to the level of expression, the tissue type(s) where expression takes place and/or developmental stage of expression. A number of heterologous regulatory sequences (e.g., promoters and enhancers) are available for controlling the expression of a HIO30.5 nucleic acid. These include constitutive, inducible and regulatable promoters, as well as promoters and enhancers that control expression in a tissue- or temporal-specific manner. Exemplary constitutive promoters include the raspberry E4 promoter (U.S. Patent Nos. 5,783,393 and 5,783,394), the 35S CaMV (Jones JD *et al*, 1992), the CsVMV promoter (Verdaguer B *et al.*, 1998) and the melon actin promoter (published PCT application WO0056863). Exemplary tissue-specific promoters include the tomato E4 and E8 promoters (U.S. Patent No. 5,859,330) and the tomato 2AII gene promoter (Van Haaren MJJ *et al.*, 1993).

[0047] In one preferred embodiment, HIO30.5 expression is under control of regulatory sequences from genes whose expression is associated with early seed and/or embryo development. Legume genes whose promoters are associated with early seed and embryo development include *V. faba legumin* (Baumlein et al., 1991, Mol Gen Genet 225:121-8; Baumlein et al., 1992, Plant J 2:233-9), *V. faba usp* (Fiedler et al., 1993, Plant Mol Biol 22:669-79), pea *convicilin* (Bown et al., 1988, Biochem J 251:717-26), pea *lectin* (dePater et al., 1993, Plant Cell 5:877-86), *P. vulgaris beta phaseolin* (Bustos et al., 1991, EMBO J 10:1469-79), *P. vulgaris DLEC2* and *PHS* [beta] (Bobb et al, 1997, Nucleic Acids Res 25:641-7), and soybean *beta-Conglycinin*, 7S storage protein (Chamberland et al., 1992, Plant Mol Biol 19:937-49). Cereal genes whose promoters are associated with early seed

and embryo development include rice *glutelin* ("GluA-3," Yoshihara and Takaiwa, 1996, Plant Cell Physiol 37:107-11; "GluB-1," Takaiwa et al., 1996, Plant Mol Biol 30:1207-21; Washida et al., 1999, Plant Mol Biol 40:1-12; "Gt3," Leisy et al., 1990, Plant Mol Biol 14:41-50), rice *prolamin* (Zhou & Fan, 1993, Transgenic Res 2:141-6), wheat *prolamin* (Hammond-Kosack et al., 1993, EMBO J 12:545-54), maize *zein* (Z4, Matzke et al., 1990, Plant Mol Biol 14:323-32), and barley *B-hordeins* (Entwistle et al., 1991, Plant Mol Biol 17:1217-31). Other genes whose promoters are associated with early seed and embryo development include oil palm GLO7A (7S globulin, Morcillo et al., 2001, Physiol Plant 112:233-243), *Brassica napus napin*, 2S storage protein, and napA gene (Josefsson et al., 1987, J Biol Chem 262:12196-201; Stalberg et al., 1993, Plant Mol Biol 1993 23:671-83; Ellerstrom et al., 1996, Plant Mol Biol 32:1019-27), *Brassica napus oleosin* (Keddie et al., 1994, Plant Mol Biol 24:327-40), *Arabidopsis oleosin* (Plant et al., 1994, Plant Mol Biol 25:193-205), *Arabidopsis* FAE1 (Rossak et al., 2001, Plant Mol Biol 46:717-25), *Canavalia gladiata* conA (Yamamoto et al., 1995, Plant Mol Biol 27:729-41), and *Catharanthus roseus* strictosidine synthase (Str, Ouwerkerk and Memelink, 1999, Mol Gen Genet 261:635-43). In another preferred embodiment, regulatory sequences from genes expressed during oil biosynthesis are used (see, e.g., US Pat No: 5,952, 544). Alternative promoters are from plant storage protein genes (Bevan et al, 1993, Philos Trans R Soc Lond B Biol Sci 342:209-15).

[0048] In yet another aspect, in some cases it may be desirable to inhibit the expression of endogenous HIO30.5 in a host cell. Exemplary methods for practicing this aspect of the invention include, but are not limited to antisense suppression (Smith, *et al.*,1988; van der Krol et al., 1988); co-suppression (Napoli, *et al.*, 1990); ribozymes (PCT Publication WO 97/10328); and combinations of sense and antisense (Waterhouse, *et al.*, 1998). Methods for the suppression of endogenous sequences in a host cell typically employ the transcription or transcription and translation of at least a portion of the sequence to be suppressed. Such sequences may be homologous to coding as well as non-coding regions of the endogenous sequence. Antisense inhibition may use the entire cDNA sequence (Sheehy et al., 1988), a partial cDNA sequence including fragments of 5' coding sequence, (Cannon et al., 1990), or 3' non-coding sequences (Ch'ng et al., 1989). Cosuppression techniques may use the entire cDNA sequence (Napoli et al., 1990; van der Krol et al., 1990), or a partial cDNA sequence (Smith et al., (1990).

[0049] Standard molecular and genetic tests may be performed to further analyze the association between a gene and an observed phenotype. Exemplary techniques are described below.

1. DNA/RNA analysis

[0050] The stage- and tissue-specific gene expression patterns in mutant versus wild-type lines may be determined, for instance, by in situ hybridization. Analysis of the methylation status of the gene, especially flanking regulatory regions, may be performed. Other suitable techniques include overexpression, ectopic expression, expression in other plant species and gene knock-out (reverse genetics, targeted knock-out, viral induced gene silencing [VIGS, see Baulcombe D, 1999]).

[0051] In a preferred application expression profiling, generally by microarray analysis, is used to simultaneously measure differences or induced changes in the expression of many different genes. Techniques for microarray analysis are well known in the art (Schena M *et al.*, Science (1995) 270:467-470; Baldwin D *et al.*, 1999; Dangond F, Physiol Genomics (2000) 2:53-58; van Hal NL *et al.*, J Biotechnol (2000) 78:271-280; Richmond T and Somerville S, Curr Opin Plant Biol (2000) 3:108-116). Expression profiling of individual tagged lines may be performed. Such analysis can identify other genes that are coordinately regulated as a consequence of the overexpression of the gene of interest, which may help to place an unknown gene in a particular pathway.

2. Gene Product Analysis

[0052] Analysis of gene products may include recombinant protein expression, antisera production, immunolocalization, biochemical assays for catalytic or other activity, analysis of phosphorylation status, and analysis of interaction with other proteins via yeast two-hybrid assays.

3. Pathway Analysis

[0053] Pathway analysis may include placing a gene or gene product within a particular biochemical, metabolic or signaling pathway based on its mis-expression phenotype or by sequence homology with related genes. Alternatively, analysis may comprise genetic crosses with wild-type lines and other mutant lines (creating double mutants) to order the gene in a pathway, or determining the effect of a mutation on expression of downstream "reporter" genes in a pathway.

Generation of Mutated Plants with an Altered Oil Content Phenotype

[0054] The invention further provides a method of identifying non-transgenic plants that have mutations in or an allele of endogenous HIO30.5 that confer a HIO30.5 phenotype to these plants and their progeny. In one method, called "TILLING" (for targeting induced local lesions in genomes), mutations are induced in the seed of a plant of interest, for example, using EMS treatment. The resulting plants are grown and self-fertilized, and the

progeny are used to prepare DNA samples. HIO30.5-specific PCR is used to identify whether a mutated plant has a HIO30.5 mutation. Plants having HIO30.5 mutations may then be tested for altered oil content, or alternatively, plants may be tested for altered oil content, and then HIO30.5 -specific PCR is used to determine whether a plant having altered oil content has a mutated HIO30.5 gene. TILLING can identify mutations that may alter the expression of specific genes or the activity of proteins encoded by these genes (see Colbert et al (2001) Plant Physiol 126:480-484; McCallum et al (2000) Nature Biotechnology 18:455-457).

[0055] In another method, a candidate gene/Quantitative Trait Locus (QTLs) approach can be used in a marker-assisted breeding program to identify alleles of or mutations in the HIO30.5 gene or orthologs of HIO30.5 that may confer altered oil content (see Bert et al., Theor Appl Genet. 2003 Jun;107(1):181-9; and Lionneton et al, Genome. 2002 Dec;45(6):1203-15). Thus, in a further aspect of the invention, a HIO30.5 nucleic acid is used to identify whether a plant having altered oil content has a mutation in endogenous HIO30.5 or has a particular allele that causes altered oil content.

[0056] While the invention has been described with reference to specific methods and embodiments, it will be appreciated that various modifications and changes may be made without departing from the invention. All publications cited herein are expressly incorporated herein by reference for the purpose of describing and disclosing compositions and methodologies that might be used in connection with the invention. All cited patents, patent applications, and sequence information in referenced public databases are also incorporated by reference.

EXAMPLES

EXAMPLE 1

Generation of Plants with a HIO30.5 Phenotype by Transformation with an Activation Tagging Construct

[0057] Mutants were generated using the activation tagging "ACTTAG" vector, pSKI015 (GI#6537289; Weigel D *et al.*, 2000). Standard methods were used for the generation of *Arabidopsis* transgenic plants, and were essentially as described in published application PCT WO0183697. Briefly, T0 *Arabidopsis* (Col-0) plants were transformed with *Agrobacterium* carrying the pSKI015 vector, which comprises T-DNA derived from the *Agrobacterium* Ti plasmid, an herbicide resistance selectable marker gene, and the 4X

CaMV 35S enhancer element. Transgenic plants were selected at the T1 generation based on herbicide resistance. T2 seed was collected from T1 plants and stored in an indexed collection, and a portion of the T2 seed was accessed for the screen.

[0058] Quantitative determination of seed fatty acid content was performed using the following methods. A sample of 15 to 20 T2 seeds from each line tested, which generally contained homozygous insertion, homozygous wild-type, and heterozygous genotypes in a standard 1:1:2 ratio, was massed on UMT-2 ultra-microbalance (Mettler-Toledo Co., Ohio, USA) and then transferred to a glass extraction vial. Lipids were extracted from the seeds and trans-esterified in 500 ul 2.5% H_2SO_4 in MeOH for 3 hours at 80 °C, following the method of Browse et al. (Biochem J 235:25-31, 1986) with modifications. A known amount of heptadecanoic acid was included in the reaction as an internal standard. 750 ul of water and 400 ul of hexane were added to each vial, which was then shaken vigorously and allowed to phase separate. Reaction vials were loaded directly onto GC for analysis and the upper hexane phase was sampled by the autosampler. Gas chromatography with Flame Ionization detection was used to separate and quantify the fatty acid methyl esters. Agilent 6890 Plus GC's were used for separation with Agilent Innowax columns (30m x 0.25mm ID, 250um film thickness). The carrier gas was Hydrogen at a constant flow of 2.5 ml/ minute. 1ul of sample was injected in splitless mode (inlet temperature 220°C, Purge flow 15ml/min at 1 minute). The oven was programmed for an initial temperature of 105°C, initial time 0.5 minutes, followed by a ramp of 60°C per minute to 175°C, a 40°C /minute ramp to 260°C with a final hold time of 2 minutes. Detection was by Flame Ionization (Temperature 275°C, Fuel flow 30.0 ml/min, Oxidizer 400.0 ml/min). Instrument control and data collection and analysis were monitored using the Millennium Chromatography Management System (Version 3.2, Waters Corporation, Milford, MA). Integration and quantification were performed automatically, but all analyses were subsequently examined manually to verify correct peak identification and acceptable signal to noise ratio before inclusion of the derived results in the study.

[0059] The ACTTAG line designated W000086431 was identified as having a high oil phenotype. Specifically, oil constituted 34.8% of seed mass (w/w) compared to an average oil content of 28.7% of other ACTTAG lines grown and analyzed in the same conditions (i.e. reference lines). Reanalysis of the same seed was performed in triplicate. Oil constituted 32.1% of seed mass, confirming an increase in oil content relative to reference.

EXAMPLE 2

Characterization of the T-DNA Insertion in Plants Exhibiting the Altered Oil Content Phenotype.

[0060] We performed standard molecular analyses, essentially as described in patent application PCT WO0183697, to determine the site of the T-DNA insertion associated with the altered oil content phenotype. Briefly, genomic DNA was extracted from plants exhibiting the altered oil content phenotype. PCR, using primers specific to the pSKI015 vector, confirmed the presence of the 35S enhancer in plants from line W000086431, and Southern blot analysis verified the genomic integration of the ACTTAG T-DNA.

[0061] Plasmid rescue and inverse PCR were used to recover genomic DNA flanking the T-DNA insertion, which was then subjected to sequence analysis using a basic BLASTN search of the Arabidopsis Genomic DNA TAIR database (available at the publicly available Arabidopsis Information Resource website). The W000086431 line has T-DNA inserted at three distinct loci.

[0062] To determine which insertion causes the high seed oil phenotype co-segregation of the high seed oil phenotype and the presence of the T-DNA was tested. Eighteen T2 plants were grown to maturity and seed harvested from these plants was used to determine the seed oil phenotype. The seed oil content from these was determined as described in Example 1. The genotype of the T2 seed was inferred by analyzing the T3 seed for the presence or absence of the T-DNA at the site of the insertion by PCR using primers that are specific to the corresponding genomic region and the T-DNA. The results show that the loci 2 and 3 were tightly linked. Furthermore, the average oil content of T3 seed containing the T-DNA insert at loci 2 and 3 was higher than those families lacking the insert at these loci. T2 individuals homozygous for loci 2 and 3 produced seed with an oil content of 115.4% of the reference, T2 individuals hemizygous for these loci produced seed with an oil content of 118.4% of the reference while T2 individuals lacking the T-DNA at these loci had an average oil content of 105% of a reference sample of seed from wild-type Col-0 plants. Because the homozygotes and hemizygotes for the high oil loci display a similar increase in oil content, we conclude that loci 2 and 3 are linked with the high oil phenotype and the phenotype is caused by a dominant mutation. By contrast, the average oil content of T3 families containing the T-DNA insert at locus 1 was lower than

or about the same as those lacking the insert at the corresponding locus. We concluded that locus 1 is not linked to the high oil phenotype.

[0063] Sequence analysis revealed that the start codon of the nucleotide sequence presented as SEQ ID NO: 1, which we designated HIO30.5, was about 1 kb 5' of the upstream border of the T-DNA insert at locus 3.

EXAMPLE 3

Analysis of *Arabidopsis* HIO30.5 Sequence

[0064] Sequence analyses were performed with BLAST (Altschul *et al.*, 1990, J. Mol. Biol. 215:403-410), PFAM (Bateman *et al.*, 1999, Nucleic Acids Res 27:260-262), PSORT (Nakai K, and Horton P, 1999,Trends Biochem Sci 24:34-6), and/or CLUSTAL (Thompson JD et al, 1994, Nucleic Acids Res 22:4673-4680).

[0065] BLASTN of SEQ ID NO:1 against GenBank sequences returned the candidate gene itself (plus its genomic DNA entry) and a putative Arabidopsis paralogue, At2g34900 (2 entries). The At3g52280 candidate gene is not represented by a cDNA clone nor are there any Arabidopsis ESTs associated with this gene:

```
   Sequences producing High-scoring Segment Pairs:                Score  P(N)
N
   gi|18409525|ref|NM_115088.1|  Arabidopsis  thaliana chromos...  5550  5.6e-
244
   gi|6434227|emb|AL132972.1|ATT25B15  Arabidopsis  thaliana D...  1155  1.0e-
175
   gi|18403687|ref|NM_129043.1|  Arabidopsis  thaliana chromos...   609  9.0e-19
1
   gi|20197115|gb|AC004238.3|  Arabidopsis  thaliana chromosom...   341  1.7e-05
2
```

[0066] BLASTP against amino acids in GenBank returned similar results, and identified 3 additional putative Arabidopsis homologues encoded by genes At2g34900 At5g10550 & At5g65630,having 39%, 13%, and 14% sequence identities, respectively, with SEQ ID NO:2. A predicted protein from rice (GI#7523514) was also identified, having 15% sequence identity with SEQ ID NO:2. The top 10 results are listed below and include human and fly genes encoding transcription factors that have been associated with controlling gene expression related to developmental patterns.

```
   Sequences producing High-scoring Segment Pairs:                Score  P(N)
N
   gi|15231174|ref|NP_190796.1|  unknown  protein; protein id:...  1570  1.1e-
160
   gi|15226857|ref|NP_181036.1|  putative  RING3 protein; prot...   606  1.5e-58
1
```

gi|15238195|ref|NP_196617.1| bromodomain protein - like; ...	239	1.0e-26	
2			
gi|7523514|dbj|BAA94242.1| Similar to Arabidopsis thalian...	278	3.4e-23	
1			
gi|15239091|ref|NP_201366.1| putative protein; protein id...	212	1.2e-21	
2			
gi|7657218|ref|NP_055114.1| bromodomain-containing protei...	215	7.9e-21	
2			
gi|3184498|gb|AAC27978.1| R31546_1 [Homo sapiens]	215	8.2e-21	
2			
gi|120558|sp|P13709|FSH_DROME FEMALE STERILE HOMEOTIC PRO...	216	1.0e-20	
2			
gi|24640482|ref|NP_511078.2| CG2252-PB [Drosophila melano...	216	1.0e-20	
2			
gi|1588281|prf||2208296A RING3 protein	255	1.2e-20	
1			

[0067] BLASTN of SEQ ID NO:1 against GenBank sequences identified the following Arabidopsis sequences with % identity to SEQ ID NO:1 indicated in parentheses: At2g34900 GI#15226857 (39%); At5g10550 GI#15238195 (13%) and At5g65630 GI#15239091 (14%).

[0068] BLASTN against ESTs returned the following ESTs from non-Arabidopsis plant species with % identity to SEQ ID NO:2 indicated in parentheses: potato (Solanum tuberosum) GI#s 13608255 (58%), and 21921887 (30%); cotton (Gossypium sp) GI#5050723 (61%); soy (Glycine max) GI#9835128 (30%) Brassica rapa GI#1019482 (51%); and corn (Zea mays) GI#24765979 (34%).

[0069] SEQ ID NOs 3 and 4 are sequences from tomato and wheat, respectively, that represent returned ESTs that were assembled into the least number of contigs. Where contigs represent partial coding regions, the entire cDNA sequence can be determined by someone skilled in molecular biology techniques.

[0070] SEQ ID NO:3 is from tomato (Lycoprsicon esculentum), and is a contig of the following sequences: GI#s 13778307, 10804738, 9504093, 6985312, 4381395. It has 58% identity with SEQ ID NO:1.

[0071] SEQ ID NO:4 is from wheat (Triticum aestivum) and is a contig of GI#s 21837181, 19950212, 20123686, 20547632, 19957908, and 20299669. It has 22% identity with SEQ ID NO:1.

[0072] PFam analysis detected a bromodomain (PF00439) from residues 87 to 200 of SEQ ID NO:2. In agreement with the predicted function as a transcription factor,

PSORT2 analysis predicts that the At3g52280 gene product is nuclear (60%). This gene may play a regulatory role in the expression of genes involved in fatty acid metabolism or related pathways.

EXAMPLE 4

Confirmation of Phenotype/Genotype Association

[0073] RT-PCR analysis showed that the HIO30.5 gene was over-expressed in plants from the line displaying the HIO30.5 phenotype. Specifically, RNA was extracted from rosette leaves and/or siliques of plants exhibiting the HIO30.5 phenotype collected at a variety of developmental stages and pooled. RT-PCR was performed using primers specific to the sequence presented as SEQ ID NO:1, to other predicted genes in the vicinity of the T-DNA insertion, and to a constitutively expressed actin gene (positive control). The results showed that plants displaying the HIO30.5 phenotype over-expressed the mRNA for the HIO30.5 gene, indicating the enhanced expression of the HIO30.5 gene is correlated with the HIO30.5 phenotype.

[0074] The dominant inheritance pattern of the HIO30.5 phenotype is confirmed through genetic analysis. In general, genetic analysis involves the production and analysis of F1 hybrids. Typically, F1 crosses are carried out by collecting pollen from T2 plants, which is used to pollinate wild type plants. Such crosses are carried out by taking about 4 flowers from each selected individual plants, and using the T2 flower as the male pollen donor and flowers of the wild type plants as the female. 4-5 crosses are done for an individual of interest. Seed formed from crosses of the same individual are pooled, planted and grown to maturity as F1 hybrids.

EXAMPLE 5

Recapitulation of the High Oil Phenotype

[0075] To confirm whether over-expression of At3g52280 causes a high seed oil phenotype, oil content in seeds from transgenic plants over-expressing this gene was compared with oil content in seeds from non-transgenic control plants. To do this, At3g52280 was cloned into a plant transformation vector behind the seed specific PRU promoter and transformed into Arabidopsis plants using the floral dip method. The plant transformation vector contains the nptII gene, which provides resistance to kanamyacin, and serves as a selectable marker. Seed from the transformed plants were plated on agar medium containing kanamycin. After 7 days, transgenic plants were identified as healthy green plants and transplanted to soil. Non-transgenic control plants were germinated on agar medium, allowed to grow for 7 days and then transplanted to soil. Twenty-two transgenic seedlings and 10 non-transgenic control plants were transplanted to random

positions in the same 32 cell flat. The plants were grown to maturity, allowed to self-fertilize and set seed. Seed was harvested from each plant and its oil content estimated by Near Infrared (NIR) Spectroscopy using methods described below.

[0076] NIR infrared spectra were captured using a Bruker 22 N/F near infrared spectrometer. Bruker Software was used to estimate total seed oil and total seed protein content using NIR data from the samples and reference methods according to the manufacturer's instructions. An oil content predicting calibration was developed following the general method of AOCS Procedure Am1-92, Official Methods and Recommended Practices of the American Oil Chemists Society, 5th Ed., AOCS, Champaign Ill). The calibration allowing NIR predictions of Crude Oil Crude Oil ASE (Ren Oil, Accelerated Solvent Extraction) was developed.

[0077] The effect of over-expression of At3g52280 on seed oil has been tested in two experiments. In both experiments, the plants over-expressing At3g52280 had higher seed oil content than the control plants grown in the same flat. Across the experiments, the average seed oil content of plants over-expressing At3g52280 was 4% greater than the untransformed controls. The in seed oil content in plants over-expressing At3g52280 was significantly greater than non-transgenic control plants (two-way ANOVA; P = 0.0208), see Table 1.

TABLE 1

Experiment	Plant ID	Transgene	Predicted average	Relative value average
1	DX04624001	PRU::HIO30.5	34.7317	105.1313
1	DX04624002	PRU::HIO30.5	35.3328	106.9509
1	DX04624003	PRU::HIO30.5	32.2034	97.4782
1	DX04624004	PRU::HIO30.5	35.2082	106.5738
1	DX04624005	PRU::HIO30.5	36.3858	110.1381
1	DX04624006	PRU::HIO30.5	35.3761	107.0819
1	DX04624007	PRU::HIO30.5	34.4489	104.2752
1	DX04624008	PRU::HIO30.5	35.1352	106.3526
1	DX04624009	PRU::HIO30.5	31.4731	95.2678
1	DX04624010	PRU::HIO30.5	34.1334	103.3205
1	DX04624011	PRU::HIO30.5	33.6363	101.8156
1	DX04624012	PRU::HIO30.5	32.135	97.2713
1	DX04624013	PRU::HIO30.5	33.0094	99.918
1	DX04624014	PRU::HIO30.5	32.8502	99.4363
1	DX04624015	PRU::HIO30.5	35.2295	106.6383

1	DX04624016	PRU::HIO30.5	32.6427	98.808
1	DX04624017	PRU::HIO30.5	33.6761	101.936
1	DX04624018	PRU::HIO30.5	32.826	99.3629
1	DX04624019	PRU::HIO30.5	31.7287	96.0416
1	DX04624020	PRU::HIO30.5	30.4119	92.0554
1	DX04624021	PRU::HIO30.5	32.6628	98.8688
1	DX04624022	PRU::HIO30.5	32.7279	99.066
1	DX04642001	None	31.8656	96.4559
1	DX04642002	None	36.3384	109.9948
1	DX04642003	None	31.6187	95.7083
1	DX04642004	None	32.7062	99.0003
1	DX04642005	None	35.1348	106.3515
1	DX04642006	None	34.2222	103.5891
1	DX04642007	None	34.5412	104.5548
1	DX04642008	None	32.7261	99.0604
1	DX04642009	None	28.0916	85.0321
1	DX04642010	None	33.12	100.2529
2	DX06812001	PRU::HIO30.5	29.234	104.5474
2	DX06812002	PRU::HIO30.5	28.1495	100.6689
2	DX06812004	PRU::HIO30.5	28.9208	103.4271
2	DX06812005	PRU::HIO30.5	32.4947	116.2083
2	DX06812006	PRU::HIO30.5	27.9377	99.9113
2	DX06812007	PRU::HIO30.5	28.9667	103.5916
2	DX06812008	PRU::HIO30.5	25.6801	91.8377
2	DX06812009	PRU::HIO30.5	27.4305	98.0976
2	DX06812010	PRU::HIO30.5	28.2639	101.0781
2	DX06812011	PRU::HIO30.5	30.168	107.8874
2	DX06812012	PRU::HIO30.5	32.4399	116.0125
2	DX06812013	PRU::HIO30.5	28.9043	103.3685
2	DX06812015	PRU::HIO30.5	32.1043	114.8121
2	DX06812016	PRU::HIO30.5	31.3237	112.0206
2	DX06812017	PRU::HIO30.5	30.9057	110.5259
2	DX06812018	PRU::HIO30.5	32.561	116.4454
2	DX06812019	PRU::HIO30.5	30.4487	108.8914
2	DX06812020	PRU::HIO30.5	31.5006	112.6532
2	DX06812021	PRU::HIO30.5	28.7262	102.7315
2	DX06812022	PRU::HIO30.5	30.3064	108.3826
2	DX06794001	None	27.007	96.5832
2	DX06794002	None	25.7445	92.0681
2	DX06794003	None	25.7293	92.0138
2	DX06794004	None	27.9487	99.951
2	DX06794005	None	28.8503	103.1751
2	DX06794006	None	27.7921	99.391
2	DX06794007	None	30.0356	107.414
2	DX06794008	None	29.9573	107.1342
2	DX06794009	None	27.1836	97.2146
2	DX06794010	None	29.376	105.0551

REFERENCES

Altschul, S.F. *et al.*, *J. Mol. Biol.* 215:403-410, 1990.

Altschul, S.F. *et al.*, *Nucleic Acids Res.* 25:3389-3402, 1997.

Ausubel FM *et al.* Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y., 1993.

Baldwin D *et al.*, *Cur Opin Plant Biol.* 2(2):96-103, 1999.

Bateman *et al.*, 1999, Nucleic Acids Res 27:260-262.

Baulcombe D, *Arch Virol Suppl* 15:189-201, 1999.

Bert et al., Theor Appl Genet. 2003 Jun;107(1):181-9.

Cannon et al., Plant Molec. Biol. (1990) 15:39-47.

Ch'ng et al., Proc. Natl. Acad. Sci. USA (1989) 86:10006-10010

Christensen S *et al.*, 9th International Conference on *Arabidopsis* Research. Univ. of Wisconsin-Madison, June 24-28, 1998. Abstract 165.

Christou et al., Proc. Natl. Acad. Sci USA (1989) 86:7500-7504.

Colbert et al (2001) Plant Physiol 126:480-484.

Cough, SJ and Bent, AF, *the Plant Journal* 16(6): 735-743, 1998.

De Block et al., Plant Physiol. (1989) 91:694-701.

Dieffenbach C and Dveksler G (Eds.) PCR Primer: A Laboratory Manual, Cold Spring Harbor Laboratory Press, NY, 1989.

Everett et al., Bio/Technology (1987) 5:1201

Feldmann *et al.*, *Science* 243: 1351-1354, 1989.

Focks N and Benning C, Plant Physiol 118:91-101, 1998.

Fridborg I *et al.*, *Plant Cell* 11: 1019-1032, 1999.

Geest AH and Hall TC, *Plant Mol Biol* 32(4):579-88, 1996.

Gelvin, S. B., Schilperoort, R. A., Varma, D. P. S., eds. Plant Molecular Biology Manual 1990.

Glick, BR and Thompson, JE, Eds. METHODS IN PLANT MOLECULAR BIOLOGY AND BIOTECHNOLOGY, p. 213-221, CRC Press, 1993.

Harlow E and Lane D, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1988, New York.

Harlow E and Lane D, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1999, New York

Hayashi H *et al.*, *Science* 258: 1350-1353, 1992.

Jako, et al., Plant Physiology 126(2):861-74, 2001.

James DW and Dooner HK (1990) Theor Appl Genet 80, 241-245.

Jensen, L.G., *et al.*, *Proc. Natl. Acad. Sci. USA* 93:3487-3491, 1996.

Jones JD *et al*, Transgenic Res 1:285-297 1992.

Kardailsky I *et al.*, *Science* 286: 1962-1965, 1999.

Katavic, et al., Plant Physiology, 108(1):399-409, 1995.

Kline et al., Nature (1987) 327:70.

Kunkel TA *et al.*, *Methods Enzymol.* 204:125-39, 1991.

Lemieux B, et al. 1990, Theor Appl Genet 80, 234-240.

Lionneton et al, Genome. 2002 Dec;45(6):1203-15).

McCallum et al (2000) Nature Biotechnology 18:455-457).

Nakamura Y et al, 1999, Nucleic Acids Res 27:292.

Napoli, *et al, Plant Cell* 2:279-289, 1990.

Omirulleh *et al.*, Plant Mol Biol. 21(3):415-28, 1993.

Sambrook *et al.* Molecular Cloning: A Laboratory Manual (Second Edition), Cold Spring Harbor Press, Plainview, N.Y., 1989.

Schaffer R, *et al.*, *Cell* 93: 1219-1229, 1998.

Sheehy et al., Proc. Natl. Acad. Sci. USA (1988) 85:8805-8809.

Smith, *et al.*, *Nature* 334:724-726, 1988.

Smith et al., Mol. Gen. Genetics (1990) 224:477-481.

Thompson JD et al, 1994, Nucleic Acids Res 22:4673-4680.

van der Krol et al., Biotechniques (1988) 6:958-976.

van der Krol et al., The Plant Cell (1990) 2:291-299.

Van Haaren MJJ *et al.*, Plant Mol Bio 21:625-640, 1993.

Verdaguer B *et al.*, Plant Mol Biol 37:1055-1067, 1998.

Waterhouse, *et al., Proc. Natl. Acad. Sci. USA* 95:13959-13964, 1998.

Weigel D, *et al.*, *Plant Physiology*, 122:1003-1013, 2000.

Wilson K *et al.*, *Plant Cell* 8: 659-671, 1996.

Yadav NS et al. (1993) Plant Physiol 103, 467-476.